ABSTRACT

Data classification using stochastic key feature generation includes obtaining a set of training data having associated summaries. The set of training data and associated summaries are used to generate a key feature generation model. Another set of training data having associated categories is also obtained, and the key feature generation model is used to map this other set of training data to a set of vectors. A data classifier is then trained based on the set of vectors and the associated categories.